217898
22P.

Performance Issues in Management of the Space Station Information System

Marjory J. Johnson

September, 1988
(revised January, 1989)

Research Institute for Advanced Computer Science
NASA Ames Research Center

RIACS TR 88.22

Research Institute for Advanced Computer Science

Performance Issues in Management of the Space Station Information System

Marjory J. Johnson

September, 1988
(revised January, 1989)

Research Institute for Advanced Computer Science
NASA Ames Research Center

The onboard segment of the Space Station Information System (SSIS), called the Data Management System (DMS), will consist of a Fiber Distributed Data Interface (FDDI) token-ring network. The purpose of this paper is to analyze performance of the DMS in scenarios involving two kinds of network management. In the first scenario we examine how the transmission of routine management messages impacts performance of the DMS. In the second scenario we examine techniques for ensuring low latency of real-time control messages in an emergency.

This work was supported by the National Aeronautics and Space Administration under NASA Contract NCC2-387 to the Universities Space Research Association (USRA).

PERFORMANCE ISSUES IN MANAGEMENT OF THE SPACE STATION INFORMATION SYSTEM

Marjory J. Johnson
Research Institute for Advanced Computer Science
NASA Ames Research Center
Moffett Field, California 94035

1. Introduction

Work is currently underway within the Space Station Program to determine how to apply emerging network-management standards to the Space Station Information System (SSIS), a wide-area networking system that encompasses space segments, ground segments, and space-to-ground links. Analyses of SSIS performance are also being conducted. Performance issues that need to be addressed with respect to management of the SSIS include both the impact of network management on overall performance of the network and the quality of service that can be provided for transmission of network-management messages themselves. In the first instance, the routine exchange of network-management messages over the network should not significantly degrade network performance. In the second instance, low latency of real-time control messages must be ensured. We address both issues in this paper.

In the past, performance hasn't been a critical consideration with respect to network management. Since management has generally consisted of passive monitoring of network operation, followed by off-line analysis of collected statistics by a human operator, network management has contributed little, if any, to the network load, and has required no special network services to support it. Consequently, few analyses of performance issues pertaining to network management have been reported in the literature. The few that exist are primarily discussions of how various monitoring techniques might perturb the network. For example, Engel [4] discusses the additional network load that is required for active, as opposed to passive, monitoring of a network.

The nature of network management is changing. As networks become more complex and as artificial-intelligence techniques are developed to remove the human operator from the management process [1,6,9,11], network management will become an active process that will contend for network resources like any other application. Accordingly, the quantification of performance issues pertaining to network management will become increasingly important.

The focus of this paper is on the portion of the SSIS called the Data Management System (DMS), which is the local area network on-board the Space Station. We investigate performance of the DMS in scenarios involving two different kinds of management functions. First, we examine how the transmission of routine monitoring messages impacts performance of the DMS. Then we examine possible techniques for ensuring low latency for management messages in a real-time fault-management scenario in which the DMS detects an emergency in one of the modules on-board the Space Station and attempts to correct the problem. The results contained herein were obtained by using a simulator developed at NASA Ames Research Center as a tool for Space Station developers.*

*This simulator, called LANES (Local Area Network Simulator), is a parameter-driven simulator of the FDDI (Fiber Distributed Data Interface) media-access-control protocol. LANES was developed by the Data Networks Concepts group at NASA Ames Research Center, under the direction of Terry Grant.

2. Space Station Information System (SSIS)

2.1. System Overview

The Space Station System will be a permanently manned international space facility, a constellation of spacecraft with the Space Station (i.e., the U. S. manned base) as its hub. It is scheduled to be on-orbit in its initial configuration in the mid 1990's. The system's lifetime will be approximately twenty-five years; it will continuously evolve over that period of time. In addition to the U. S. manned base, the Space Station System will include co-orbiting platforms, polar-orbiting platforms, free flyers, orbital-transfer vehicles, orbital-maneuvering vehicles, and pressurized modules provided by the European Space Agency and the National Space Development Agency of Japan. The system will serve as a scientific laboratory, as a manufacturing facility, and as a base for deep-space exploration.

Two types of payloads will be supported. Payloads that require crew interaction, primarily materials processing and life-sciences experiments, will be located in the pressurized modules. Examples of materials processing include the growth of crystals for silicon wafers, the development of ultra-pure chemicals for drugs, and the development of techniques for building structures to be used for deep-space exploration. Life-sciences experiments test man's ability to adapt to the harsh environment of space; examples include experiments to determine the impact of artificial gravity on bone formation and to determine the effects of long-term weightlessness on the heart. Payloads of an observational nature will be located primarily on the unpressurized platforms. Examples include observations of sun-spot activity and observations of weather patterns on the earth. Instruments located on the platforms will collect data to be transmitted to scientists on the ground for analysis.

The SSIS will provide end-to-end information services for the Space Station and platforms. It must handle critical functions, such as navigation, environmental control, and life-support, as well as handling the massive volumes of data that will be generated by the payloads. The space-segment elements of the SSIS, i.e., the local area networks on-board the individual spacecraft, will be connected by radio-frequency links. Gateways will connect the SSIS to the European and Japanese modules. The ground segment in the United States is a wide-area internetworking system which includes control and data-capture facilities, operations-support networks, and nation-wide internetworks to serve the scientific and research communities. The space-to-ground link will be provided by a system of communications satellites, called the Tracking and Data Relay Satellite System (TDRSS).

2.2. Data Management System and FDDI Protocol

The portion of the SSIS that is located on-board the Space Station is called the Data Management System (DMS). The current proposal calls for the DMS to be a Fiber Distributed Data Interface (FDDI) token ring. Accordingly, the analysis presented herein is an analysis of FDDI performance in the context of management of the DMS.

FDDI is an emerging American National Standards Institute (ANSI) and International Standards Organization (ISO) standard for a 100 megabit-per-second fiber-optic token ring. Since an understanding of the prioritization mechanisms of FDDI is necessary for our analysis, we include a brief discussion of the FDDI access protocol.

FDDI is a timed-token-rotation protocol; timers within each node cooperatively attempt to maintain a specified token-rotation time by using the observed network load to regulate the amount of time that an individual node may transmit. As a result of this regulation, token-rotation time is bounded. This feature enables FDDI to support two classes of service: synchronous service for applications with stringent channel-access requirements, such as packet voice and real-time control, and asynchronous service for applications which do not have such stringent channel-access requirements.

At ring initialization, nodes negotiate the value to be assigned to *T_Opr*, a parameter which specifies the expected token-rotation time. The smallest requested value is assigned to *T_Opr*, so that timing constraints of all the nodes will be satisfied. Each node is assigned a percentage of *T_Opr* for its synchronous-bandwidth allocation. A node may transmit synchronous frames for its allotted time whenever it receives the token. Asynchronous frames may be transmitted only if the load on the ring, as measured by the amount of time required for the preceding token rotation, is light enough to permit it. Hence, synchronous traffic has higher priority than asynchronous traffic. FDDI supports multiple priorities of asynchronous traffic, via a mechanism called the priority threshold. Each priority of asynchronous traffic has an associated priority threshold, which is expressed in units of time.

When the token arrives at a node, that node first transmits synchronous frames, up to its pre-assigned synchronous-bandwidth allotment, and then transmits its asynchronous frames, highest priority frames first, as long as internal timers permit. When the token arrives at a node, an internal token-holding timer is set to the amount of time remaining in a *T_Opr* time period which began with the previous arrival of the token at that node.* This timer is enabled (counting down to 0) only while the node transmits asynchronous frames; it's value at any given time represents the maximum amount of time available to that node for asynchronous transmission. Before a node is allowed to transmit a new asynchronous frame, the priority threshold associated with that frame is compared with that node's remaining token-holding time. If the remaining token-holding time is greater than the priority threshold, then the frame is transmitted. Otherwise, the token is forwarded to the next node on the ring. It follows that the lower the value of the priority threshold associated with a particular traffic class, the higher the priority of that class. The priority threshold associated with highest-priority asynchronous traffic is 0, so that a node may transmit this type of traffic if any time at all remains on its token-holding timer. We make the reasonable assumption that there are separate queues within each node for synchronous frames and for each priority of asynchronous frames. In this way lower-priority frames cannot block higher-priority frames from being transmitted.

For more detail about the FDDI media-access-control protocol see [5,10]. Several analyses of FDDI traffic classes appear in the literature. Goyal and Dias [7] compare prioritization mechanisms of FDDI and the IEEE 802.5 token ring. Dykeman and Bux [2,3] determine maximum throughput levels for each asynchronous priority class as a function of the associated priority-threshold value and the value of *T_Opr*. Johnson [8] analyzes synchronous traffic delays. The emphasis in the present paper is in determining how to utilize FDDI prioritization techniques to minimize delay of various traffic classes in typical management scenarios.

---

*The *T_Opr* time period will actually begin prior to the arrival of the token at a node if the token is running behind schedule.

2.3. Management of the DMS

NASA's space-mission networking environment presents some unique problems for network management. The remoteness of the DMS, located in space, means that human operators won't be readily available to monitor and maintain the system. The Space Station will be manned, but it is undesirable to use scarce crew time for routine maintenance. Yet, reliability of the system is critically important, because lives of the crew members may depend on its correct functioning. Extensive management of the system from earth is also undesirable because of the significant transmission delay in the space-to-ground link.

These considerations lead to two requirements for management of the DMS: the distribution of management functionality and the use of automation. Distribution of management functionality will enhance reliability and guard against the existence of a single point of failure on the network. Automation of management functionality will eliminate, or at least considerably reduce, the need for human-operator involvement. Intelligent agents, which may be expert-system based or may use other technologies, will play a major role in management of the DMS, as well as in other areas of operational control of the Space Station, especially in the evolutionary phases of the Space Station System. The interaction of these intelligent agents to coordinate their activities may have a significant impact on the DMS workload, especially with respect to requirements on message latency.

Performance may be an issue in management of the DMS in two ways. First, it is important that routine management messages have minimal impact on performance of the rest of the network. Second, when coping with an emergency, management messages themselves must receive adequate service to support intelligent agents which are involved in real-time control of critical life-support functions.

3. Management Scenarios

In this section we present scenarios involving two different kinds of network-management functionality. The ring configuration used for our analysis is presented in Table 1. The simulator we used is a detailed model of the FDDI media-access-control protocol. Delays due to processing at the upper layers of the Open Systems Interconnection (OSI) network model, either at the source or at the destination, are not factored into this study. When frames are generated, they are immediately queued for transmission at the source (i.e., there is no passage of simulated time). Each node has infinite buffer capacity, thus eliminating possible buffering delays. Hence, total frame delay measured in the simulation consists of queueing delay at the source node, transmission time, and the time required for the frame to propagate from the source node to the destination node.

Table 1. Ring Configuration

Parameter	Value
Number of Nodes	20
Distance between Nodes	30 meters
Management-frame size	100 bytes
Background-frame size	2000 bytes
Header size	40 bytes
T_Opr	20 milliseconds

3.1. Routine Management Messages

Under normal operating conditions, management of the DMS will consist primarily of performance monitoring and reporting. In this section we analyze how the transmission of routine management messages, such as messages reporting status of various local resources or values of local parameters, impacts performance of the DMS. Since management functionality will be distributed on the DMS, with each node contributing equally to the management process, we assume that each node sends and receives approximately the same amount of management information over the network. We also assume that management messages will be relatively short.

For this analysis we modeled the transmission of an average of 200 management frames per second from each node. Management traffic, including header overhead, thus accounts for approximately 5% of the capacity of the ring. This additional traffic on the network will, of course, increase message delay for the background messages. Since it is desirable that it have minimal impact on delay of background messages, routine management traffic should be assigned lower priority than background traffic. For our analysis we modeled both background traffic and management traffic as asynchronous traffic, with management traffic having lower priority. We set the priority threshold value for the management traffic equal to one-half the *T_Opr* value. This setting significantly restricts access to the network for transmission of management messages when the ring is heavily loaded. The volume of the background loading was varied, so that the impact of management traffic on the background traffic could be observed under different traffic loads. Interarrival times for both management and background frames were exponentially distributed.

Figure 1 compares average queueing delays for management frames and background frames.* Offered load in this figure refers only to the volume of background traffic, i.e., it does not include management traffic. Until the ring is loaded at approximately 85% of capacity, queueing delays for both types of traffic are approximately the same. As the offered load increases, the difference in queueing delays experienced by the two types of traffic also increases. In our experiment when the background traffic reached or exceeded the capacity of the ring, then management traffic was completely blocked from the ring. This behavior is desirable, since routine management messages are not critical to the functioning of the DMS.

Figure 2 compares average queueing delays experienced by the background traffic when it is the only traffic on the ring and when it is accompanied by the 5% level of lower-priority routine management traffic. Again, offered load refers only to the volume of background traffic. The shape of both delay graphs shows that the additional management traffic has an insignificant impact on the overall message delay until the ring is loaded at approximately 85% of capacity. It is also clear from our studies that designating management traffic to be lower priority minimizes the queueing delays experienced by background traffic when the ring is heavily loaded. Since the DMS is not expected to be heavily loaded during normal operation, we conclude that transmission of a low volume of routine management frames as lower-priority asynchronous traffic will not significantly degrade performance of the system.

---

*As indicated above, total frame delay measured in the simulation includes queueing delay at the source node, transmission time, and the time required for the frame to propagate from the source node to the destination node. Because of the difference in size between management frames and background frames, which causes a significant difference in transmission time, results herein are presented in terms of queueing delay only.

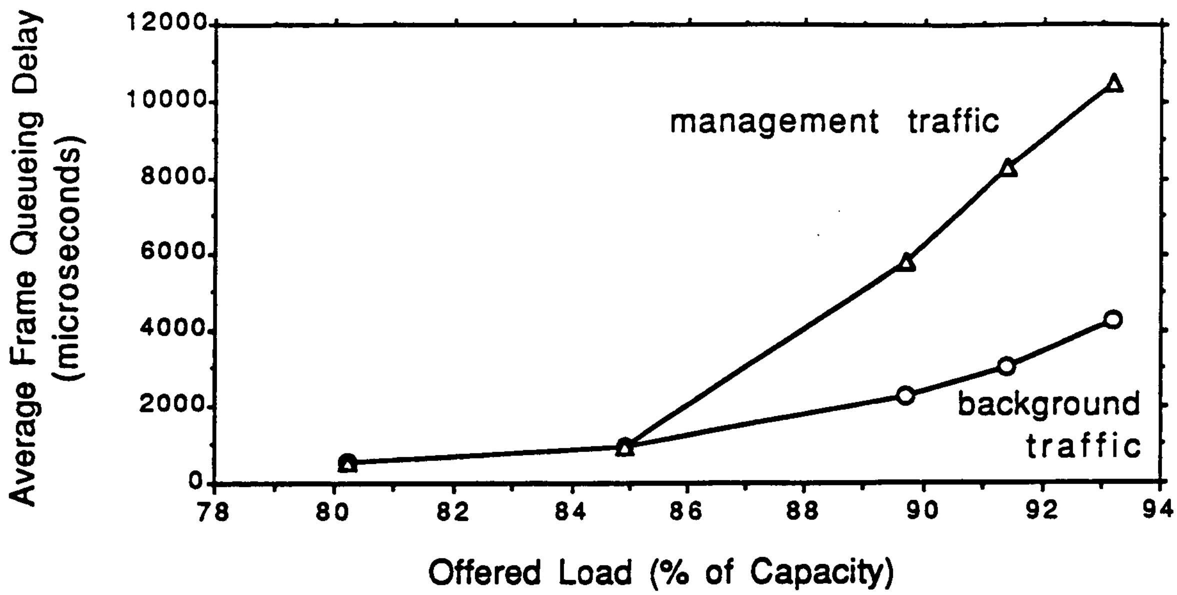

Figure 1. Comparison of Average Queueing Delays for Different Traffic Classes

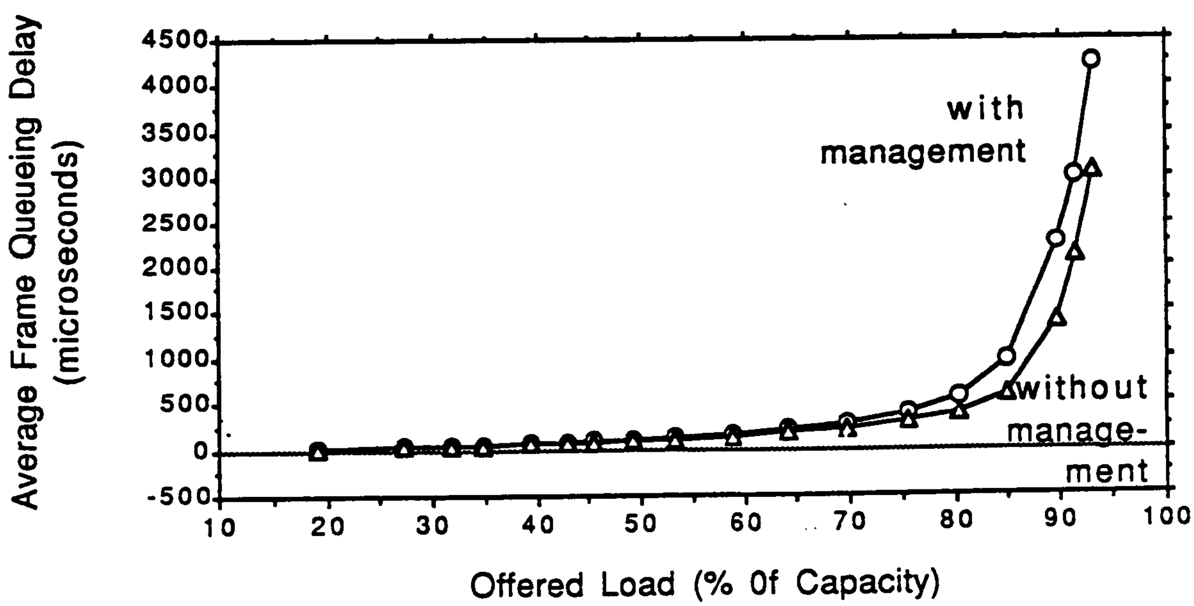

Figure 2. Comparison of Average Queueing Delays for Background Traffic, with and without Management Traffic

We also experimented with smaller-sized background frames and larger volumes of management traffic. The same basic queueing-delay patterns as presented in Figures 1 and 2 were observed. Either of these situations causes the delay curves to go to infinity at lower ring loadings.

3.2. Handling Emergencies

If a life-critical or mission-critical situation were to arise in one of the modules onboard the Space Station, the load on the DMS could increase dramatically as the system tried to identify and correct the problem in an automated fashion. Although such an emergency might not be caused by a fault within the DMS itself, the DMS would probably be heavily involved in correcting the problem. In such a situation the traffic to handle the emergency would need to be assigned highest priority, since ensuring low latency for this traffic would be critical.

Using the ring configuration presented in Table 1, changing only the *T_Opr* value, we modeled a scenario in which five nodes exchange messages to handle an emergency that involves a module supported by these nodes. These five nodes transmit short, frequent emergency messages (corresponding to the management messages described in Table 1). The remaining fifteen nodes on the ring continue transmission of background messages. The volume of the emergency messages is approximately 10 megabits per second, i.e., approximately 10% of ring capacity. We selected this figure because a single network-interface unit on the DMS is required to handle a maximum throughput of ten megabits per second. In an emergency one node might be acting as a master to coordinate activities of the other nodes, and 10 megabits per second is the maximum volume of traffic that this node would be able to handle. As in the previous scenario, the volume of background traffic is varied, so that we can observe latency of emergency traffic under different traffic loads. Also as before, all interarrival times for both emergency and management frames are exponentially distributed.

The upper bound on token-rotation time in an FDDI token ring is $2 \times T\_Opr$ [5]. Thus, the smaller the value of *T_Opr*, the more quickly the token is forced to rotate around the ring. We assigned a relatively small value of 2 milliseconds to *T_Opr* for this scenario, thus guaranteeing frequent channel access to the nodes handling the emergency. In addition, we assigned a higher priority to emergency traffic than to background traffic, to ensure that emergency traffic would receive superior service. We prioritized the traffic in two different ways, to determine the more effective way to minimize latency for the emergency traffic. The first method was to designate both emergency and background traffic to be asynchronous traffic, with emergency traffic having higher priority than the background traffic. In this case, the priority threshold for the background traffic was set to $T\_Opr/2$. The second method was to designate emergency traffic to be synchronous traffic and background traffic to be highest-priority asynchronous traffic.

Figure 3 compares average queueing delays for emergency traffic and background traffic, when asynchronous priorities are used to differentiate between the quality of service provided to the two types of traffic. A difference in queueing delay becomes noticeable when the ring is loaded at approximately 70%, and of course, is more and more significant as ring loading increases.* Even when the offered load exceeds the capacity of the ring and queueing delays for the background traffic become unbounded, the average queueing delay for emergency traffic is below 500 microseconds. Hence, this is an effective technique for minimizing latency for the emergency traffic.

---

*In this section, offered load includes both emergency traffic and background traffic.

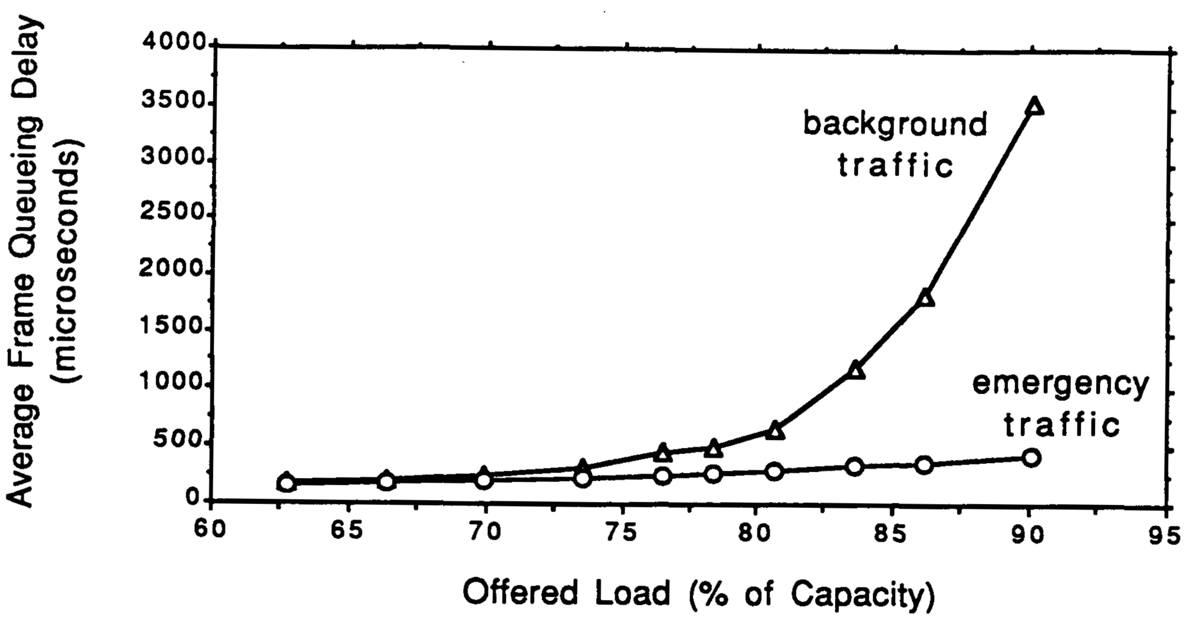

Figure 3. Comparison of Average Queueing Delays for Different Traffic Classes, Using Asynchronous Priorities to Distinguish between the Classes

Next, we compared queueing delays for emergency traffic and background traffic, when synchronous bandwidth is reserved for emergency traffic and background traffic is highest-priority asynchronous traffic. In this case the synchronous-bandwidth assignments were sufficient to accommodate all emergency traffic that might be generated within a single *T_Opr* time period. The general pattern that we observed for average queueing delay for emergency traffic as compared to average queueing delay for background traffic is similar to the pattern depicted in Figure 3. The major difference is that when synchronous bandwidth is used to assign highest priority to emergency traffic, both average and maximum queueing delays for emergency traffic are significantly higher than when both classes of traffic are asynchronous. This phenomenon is displayed in Figures 4 and 5. Such behavior may seem counter-intuitive, since the purpose of the FDDI synchronous-service class is to provide guaranteed timely access to the channel to transmit synchronous frames. However, the explanation for this behavior is

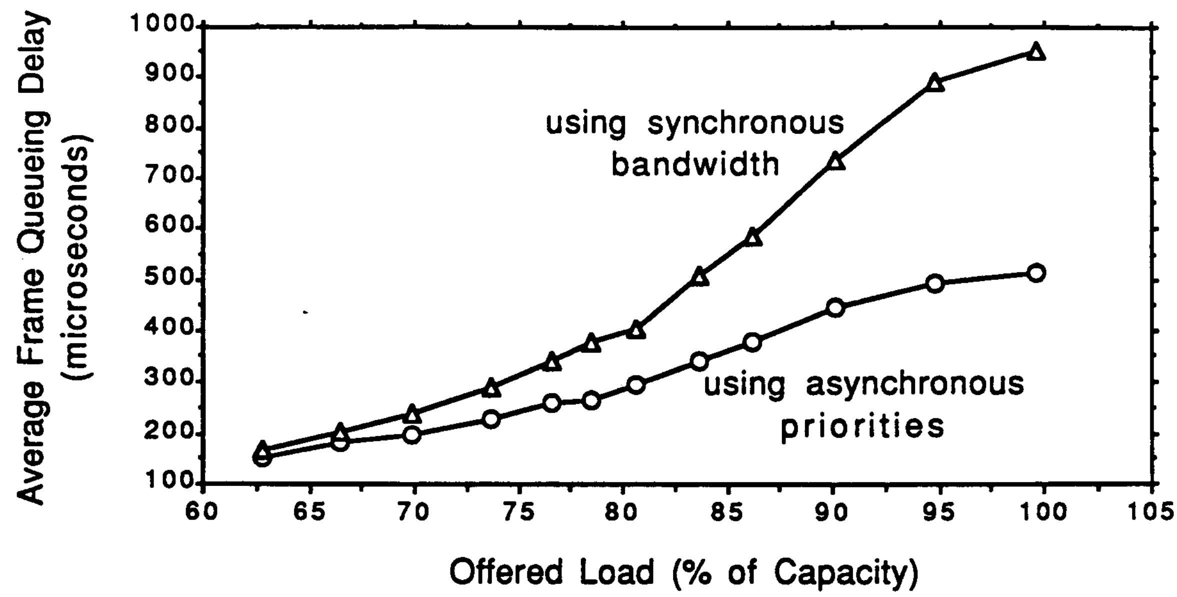

Figure 4. Comparison of Average Queueing Delays for Emergency Traffic, Using Different Prioritization Techniques

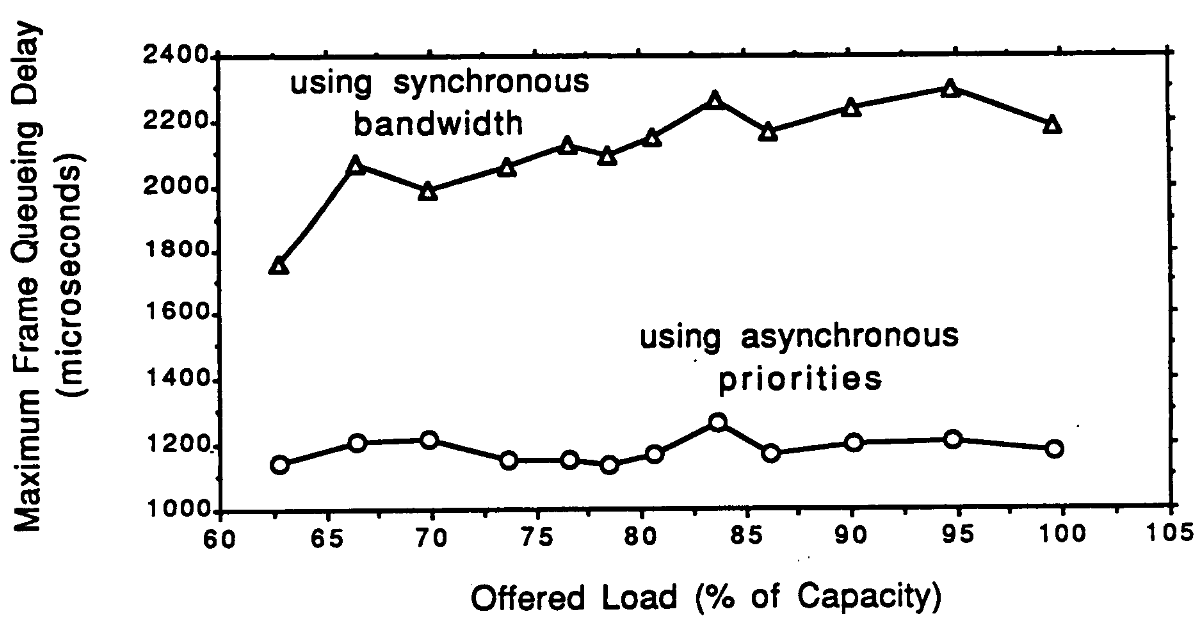

Figure 5. Comparison of Maximum Queueing Delays for Emergency Traffic, Using Different Prioritization Techniques

straightforward. Furthermore, the delay pattern exhibited in Figures 4 and 5 is actually independent of the values assigned either to *T_Opr* or to the priority threshold associated with background traffic in the case when both types of traffic are asynchronous.

The key to minimizing latency for the highest-priority (i.e., emergency) traffic is to ensure short token-rotation times. The values assigned to *T_Opr* and to the priority thresholds for asynchronous traffic both affect the speed of token rotation. That the value of *T_Opr* affects token-rotation time is obvious, because the upper bound on token-rotation time is a function of this parameter. In order to satisfy latency requirements for emergency traffic, it might be necessary to reconfigure the ring to lower the value of *T_Opr*.

The effect of priority-threshold assignments on token-rotation time is not as obvious, but is significant nevertheless. The priority threshold associated with a traffic class limits access to the ring for that class. The larger the priority threshold, the greater the restriction. This is because a node which is holding the token may transmit a pending asynchronous frame only if the time remaining on its token-holding timer exceeds the priority threshold of that frame. For a given traffic load and physical ring configuration, increasing the value of the priority threshold associated with one of the asynchronous traffic classes, while keeping all others fixed, tends to reduce the average number of frames of that particular traffic class that are transmitted per token rotation. This tends to increase the speed of token rotation, which tends to provide more frequent channel access to higher-priority traffic. The end result is that higher-priority traffic tends to experience lower queueing delays. The advantage to higher-priority traffic which results from increasing the limitation on transmission of lower-priority traffic is especially noticeable under heavy traffic loads.

The key to explaining the pattern of delays shown in Figures 4 and 5 is the handling of background traffic, rather than the handling of emergency traffic. In our scenario the priority threshold associated with background traffic was 0 when synchronous bandwidth was reserved for emergency traffic, and was *T_Opr*/2 when both traffic classes were asynchronous. As expected, based on the discussion above, we observed that the average token-rotation time was less when both traffic classes were asynchronous than when synchronous bandwidth was used for emergency traffic. It is this phenomenon that causes the difference in delays shown in Figures 4 and 5. In our scenario, the difference between the two priority-threshold values assigned to the background traffic is significant. This causes a significant difference in average token-rotation times, which in turn causes a significant difference in delays in the two situations. A larger volume of emergency traffic in our scenario would have lessened the dependence of average token-rotation time on the average number of background frames transmitted per token rotation, and hence would have lessened the difference in delays caused by the different prioritization techniques.

We conclude that the more effective way to minimize latency for emergency traffic for the DMS is to transmit both emergency traffic and background traffic as asynchronous traffic, designating higher priority for the emergency traffic. This method is also simpler to implement, for the time required for a management authority to instruct nodes to lower the asynchronous priority of their background traffic would be much less than that required to compute new synchronous-bandwidth requirements. As indicated above, it might also be necessary to reconfigure the ring to lower its *T_Opr* value when the DMS detects an emergency, in order to satisfy latency requirements for the emergency traffic. Both the changing of parameter values and reconfiguration of the ring, if necessary, would introduce delays in the network's response to the emergency. The impact of these delays will be studied experimentally using DMS testbeds that are being developed at various NASA centers.

4. Conclusions

As the Space Station evolves, operations management (including network management) will become increasingly automated. Distributed intelligent agents will work autonomously to accomplish everything from routine monitoring, to on-line analysis of system performance, to fault detection, isolation, and repair. The effectiveness of this approach to management depends not only on correctness of algorithms that might be used, but also on some performance considerations. Ideally, management should be virtually transparent to a system user when the system is working well, but should be the highest-priority task in emergencies.

In this paper we examined performance issues pertaining to management of the Data Management System of the Space Station Information System. First, we examined the impact of routine management traffic on background traffic during normal operation. Second, we compared two techniques for handling emergency traffic, to determine which would provide lower latency.

We conclude that the DMS can handle routine management traffic without significantly degrading network performance and can ensure that management traffic will experience sufficiently low latency to cope effectively with emergencies. The FDDI media-access-control protocol provides sufficient prioritization mechanisms to distinguish between the quality of service provided to management and background traffic in either situation. We found that the use of asynchronous priorities to distinguish between the two traffic classes is the more effective way to minimize latency for management traffic in emergencies.

The analysis herein is not intended to provide assistance in selecting optimal parameter settings for FDDI in specific network-management situations. Instead, it identifies some network-management performance issues that are important in the context of the Space Station, and demonstrates that the DMS would be able to handle management traffic effectively in various situations. In the future we believe that similar types of performance issues will assume significance for ground-based networks, as well as networks in space, as automation assumes a more prominent role in network management.

References

1. L. Bernstein and C. M. Yuhas, "Expert Systems in Network Management - The Second Revolution," *IEEE J. Sel. Areas in Commun.*, Vol. SAC-6, No. 5, June 1988, pp. 784-787.

2. D. Dykeman and W. Bux, "An Investigation of the FDDI Media Access Control Protocol," *Proc. EFOC/LAN '87*, Basel, Switzerland, June, 1987, IGI Europe - Information Gatekeepers, Inc., Boston, pp. 229-236.

3. D. Dykeman and W. Bux, "Analysis and Tuning of the FDDI Media-Access Control Protocol," *IEEE J. Sel. Areas in Commun.*, Vol. SAC-6, No. 6, July, 1988, pp. 997-1010.

4. T. Engel, LAN Monitoring Concepts for Management Support, *Proceedings of the Sixth European Fibre Optic Communications and Local Area Networks Conference*, Amsterdam, The Netherlands, June, 1988, IGI Europe - Division of Information Gatekeepers, Inc., pp. 446-451.

5. *FDDI Token Ring Media Access Control*, American National Standard, X3.139 - 1987.

6. M. Feridun, M. Leib, M. Nodine, and J. Ong, "ANM: Automated Network Management System," *IEEE Network*, Vol. 2, No. 2, March 1988, pp. 13-19.

7. A. Goyal and D. Dias, "Performance of priority protocols on high-speed token ring networks," *Proc. 3rd Int. Conf. on Data Communication Systems and Their Performance*, Rio de Janeiro, Brazil, June 1987, pp. 25-34.

8. M. J. Johnson, "Analysis of FDDI Synchronous Traffic Delays," *Proceedings of Systems Design and Networks Conference: Putting Local Area Networks to Work*, Santa Clara, CA, April, 1988, IEEE, pp. 65-72.

9. D. Peacocke and S. Rabie, "Knowledge-Based Maintenance in Networks," *IEEE J. Sel. Areas in Commun.*, Vol. SAC-6, No. 5, June 1988, pp. 813-818.

10. F. E. Ross, "FDDI - A Tutorial," *IEEE Commun. Mag.*, Vol. 24, no. 5, 1986, pp. 10-17.

11. R. O. Yudkin, "On Testing Communication Networks," *IEEE J. Sel. Areas in Commun.*, Vol. SAC-6, No. 5, June 1988, pp. 805-812.